

# On Transmitting Correlated Sources over a MAC

Arun Padakandla

**Abstract**—The joint source channel coding problem of transmitting a pair of correlated sources over a 2-user MAC is considered. A new concatenated coding scheme, comprising of an inner code of fixed block-length and an outer code of arbitrarily large block-length, is proposed. Its information theoretic performance is analyzed to derive a new set of sufficient conditions. An example is identified to demonstrate that the proposed coding technique can strictly outperform the current known best, which is due to Cover El Gamal and Salehi [1]. Our findings are based on Dueck’s ingenious coding technique proposed for the particular example studied in [2].

## I. INTRODUCTION

In several multi-terminal communication, cryptographic and information processing scenarios, the presence of a Gács Körner (GK) common part leads to richer strategies and enhanced throughput. For example, in quantizing [3] or communicating [1] correlated sources that are distributed, the GK common part facilitates co-ordination through a common codebook. In communicating private messages over a broadcast channel, the GK common part of the induced test channel can be specifically coded to enhance throughput. In this article, we address the scenario wherein the sources nearly, but not perfectly, have a GK common part, and we develop a strategy to exploit this specific structure of the joint pmf.

Our primary focus in this article is the scenario depicted in Fig. 1, wherein a pair  $S_1, S_2$  of correlated sources have to be communicated losslessly over a 2-user MAC. We undertake a Shannon-theoretic study of this scenario and consider the problem of characterizing sufficient conditions under which the sources can be communicated over the MAC.

This problem was addressed by Cover, El Gamal and Salehi [1], wherein a coding scheme (CES scheme) that transfers the source correlation into correlated channel inputs is proposed. The CES scheme outperforms separation, and remains to be the best known scheme for a general problem instance. While the CES scheme transfers the source correlation into channel inputs via a ‘single-letter’ scheme, i.e., the correlation between  $X_1$  and  $X_2$  at time  $t$  is based only on the correlation between the source symbols at time  $t$ , Dueck [2], through an ingenious coding scheme designed for a specific example, proves that if one allows channel inputs to extract correlation from a block of symbols, then these can be carefully designed to communicate more effectively over the MAC. Thereby, Dueck proves that the single-letter (S-L) CES scheme is sub-optimal, however his coding scheme is very specific to the example considered therein. In this work, we build on Dueck’s findings and propose a generalized coding scheme.

Arun Padakandla is with the Center for Science of Information at Purdue University.

This work was supported by NSF grants NSF CCF-0939370 and NSF CCF-1140489.

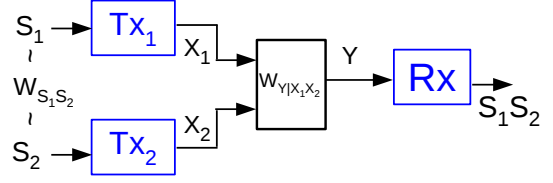


Fig. 1. Transmission of correlated sources over MAC.

Our contributions also include a performance analysis of the proposed coding scheme which leads us to characterizing a new set of sufficient conditions. In this endeavor, we are aided by recent findings of Shirani and Pradhan [4], wherein a closely related phenomenon is identified and exploited for distributed source coding (DSC).

Joint source-channel coding has been well studied, both from the perspective of fundamental limits [5], [6] and computationally feasible strategies [7], in several multi-terminal scenarios [8], [9]. Fundamental limits of performance for transmitting correlated Gaussians over Gaussian multi-user channels have been studied in [6], [10].

Following are the primary motivating questions we address in our work. Firstly, why does the flexibility of allowing channel inputs at time  $t$  to be correlated via a block of source symbols, not just those at time  $t$ , provide, in general, a more efficient technique for the problem considered herein. In other words, why is the S-L CES scheme sub-optimal? In Section II, we provide a discussion on Dueck’s findings, and explain why a S-L CES scheme, or perhaps any scheme constrained to transfer S-L correlation, is, in general, sub-optimal. Secondly, can we employ the current known S-L schemes as building blocks in a larger coding scheme and thereby transfer correlation from a block of source symbols onto channel inputs? A careful understanding of Dueck’s coding scheme suggests, as stated in Section II, that a concatenated coding scheme, wherein both the inner and outer<sup>1</sup> codes operate in a single letter fashion can accomplish this. Finally, can we characterize the performance of such a concatenated coding scheme (that is computable) and compare that to a S-L CES scheme?

Taking a closer look at Dueck’s findings in section II, we recognize why a S-L CES scheme is sub-optimal. Therein, it also becomes apparent how the specific structure of the joint pmf of  $S_1, S_2$  ‘nearly’, but not perfectly, possessing a GK common part can be exploited via a concatenated coding scheme involving S-L coding schemes. In section III, we propose this coding scheme, analyze its performance to derive a computable characterization of a set of sufficient conditions. In doing this, we employ techniques from Chaharsooghi and

<sup>1</sup>The inner code is of fixed block-length, and outer code is of arbitrarily large block length

Pradhan [4], [11] wherein a similar effort is carried out in the context of DSC. In section III, we also prove, via a simple modification of Dueck's example, that the derived sufficient conditions can be strictly less binding than those of CES [1].

The significance of our contribution is highlighted. Firstly, our study puts forth a new coding scheme for communicating correlated sources over a MAC and characterizes a new set of sufficient conditions for transmissibility. Through an example, we demonstrate that the proposed coding technique can strictly outperform the current known best. Secondly, our findings indicate that, even though a S-L scheme might not, in general be optimal for a multi-terminal info-theory problem, one might be able to provide better approximations of performance limits through computable characterizations. Thirdly, our coding scheme indicates that one might be able to approach the performance of multi-letter coding schemes by appropriately stitching together S-L coding schemes. Our findings provide a new set of tools and techniques to explore several fundamental multi-terminal information theory problems that have resisted progress for over three decades.

#### A. Preliminaries : Notation and problem statement

We supplement standard information theory notation - upper case for RVs, calligraphic for sets etc. - with the following. We let an underline denote an appropriate aggregation of related objects. For example,  $\underline{S}$  will be used to represent a pair  $S_1, S_2$  of RVs.  $\underline{S}$  will be used to denote either the pair  $S_1, S_2$  or the Cartesian product  $\mathcal{S} \times \mathcal{S}_2$ , and will be clear from context. When  $j \in \{1, 2\}$ , then  $\bar{j}$  will denote the complement index, i.e.,  $\{j, \bar{j}\} = \{1, 2\}$ . For  $m \in \mathbb{N}$ ,  $[m] := \{1, \dots, m\}$ .

$$T_\delta^n(\mathcal{U}) = \{u^n \in \mathcal{U}^n : \left| \frac{N(b|u^n)}{n} - p_U(b) \right| \leq \delta p_U(b) \forall b \in \mathcal{U}\}$$

will be our typical set. For a pmf  $p_U$  on  $\mathcal{U}$ ,  $b^* \in \mathcal{U}$  will denote a symbol with the least positive probability wrt  $p_U$ .<sup>2</sup> Boldface letters such as  $\mathbf{A}$  denote matrices. For a  $m \times l$  matrix  $\mathbf{A}$ , (i)  $\mathbf{A}(t, i)$  denotes the entry in row  $t$ , column  $i$ , (ii)  $\mathbf{A}(1 : m, i)$  denotes the  $i^{\text{th}}$  column,  $\mathbf{A}(t, 1 : l)$  denotes  $t^{\text{th}}$  row. The following words/phrases are used often and hence abbreviated. "high probability", "with high probability", "single-letter", "long Markov chain", "random variable" are abbreviated hp, whp, S-L, LMC, rv respectively.

Consider a 2-user MAC with input alphabets  $\mathcal{X}_1, \mathcal{X}_2$  and output alphabet  $\mathcal{Y}$ . Let  $\mathbb{W}_{Y|X_1 X_2}$  denote the channel transition probabilities. Let  $\underline{S} := (S_1, S_2)$  taking values over  $\underline{\mathcal{S}} := \mathcal{S}_1 \times \mathcal{S}_2$  with pmf  $\mathbb{W}_{S_1 S_2}$  denote a pair of information sources. For  $j \in [2]$ , encoder  $j$  observes  $S_j$ . The decoder aims to reconstruct  $\underline{S}$  with arbitrarily small probability of error. In this article, our objective is to characterize sufficient conditions for transmissibility of sources  $(\underline{S}, \mathbb{W}_{\underline{S}})$  over the MAC  $(\underline{\mathcal{X}}, \mathcal{Y}, \mathbb{W}_{Y|\underline{\mathcal{X}}})$ .

#### II. KEY ELEMENTS OF DUECK'S CODING SCHEME [2]

Dueck's example [2] corresponds to  $\eta = 1$  in the following simple generalization. In describing the following example, we employ Dueck's notation.

<sup>2</sup>The underlying pmf  $p_U$  will be clear from context.

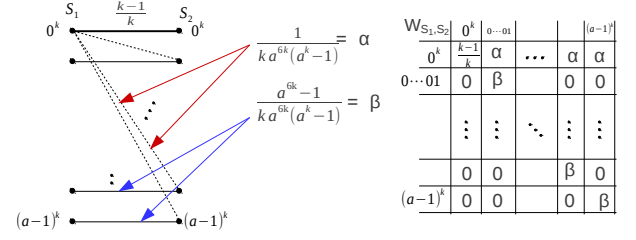


Fig. 2. On the left, the source pmf is depicted through a bipartite graph. Larger probabilities are depicted through edges with thicker lines. On the right, we depict the probability matrix.

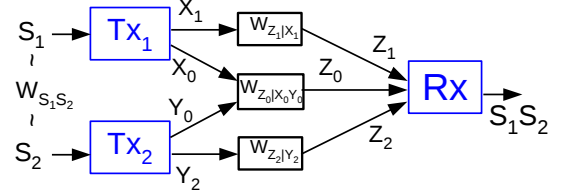


Fig. 3. Source channel setup of Example 1.

*Example 1:* Source alphabets  $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1, \dots, a-1\}^k$ . Let

$$\mathbb{W}_{S_1 S_2}(c^k, d^k) = \begin{cases} \frac{k-1}{k} & \text{if } c^k = d^k = 0^k \\ \frac{a^{\eta k} - 1}{ka^{\eta k}(a^k - 1)} & \text{if } c^k = d^k, c^k \neq 0^k, \text{ and} \\ \frac{1}{ka^{\eta k}(a^k - 1)} & \text{if } c^k = 0^k, d^k \neq 0^k \end{cases}$$

where  $\eta$  is a positive integer. Note that in the above eqn.  $c^k, d^k \in \mathcal{S}_1$  abbreviate the  $k$  'digits'  $c_1 c_2 \dots c_k$  and  $d_1 d_2 \dots d_k$  respectively. Fig. 2 depicts the source pmf with  $\eta = 6$ .

The MAC is depicted in Fig. 3 and described below. The input alphabets are  $\underline{\mathcal{X}}$  and  $\underline{\mathcal{Y}}$ , where  $\underline{\mathcal{X}} = \mathcal{X}_0 \times \mathcal{X}_1$  and  $\underline{\mathcal{Y}} = \mathcal{Y}_0 \times \mathcal{Y}_2$ . The output alphabet  $\underline{\mathcal{Z}} = \mathcal{Z}_0 \times \mathcal{Z}_1 \times \mathcal{Z}_2$ , where  $\mathcal{X}_0 = \mathcal{Y}_0 = \mathcal{Z}_0 = \{0, 1, \dots, a-1\}$ . Moreover,  $\mathbb{W}_{Z_0|\underline{\mathcal{X}}\underline{\mathcal{Y}}} = \mathbb{W}_{Z_0|X_0 Y_0}$ ,  $\mathbb{W}_{Z_1|\underline{\mathcal{X}}\underline{\mathcal{Y}}} = \mathbb{W}_{Z_1|X_1}$ ,  $\mathbb{W}_{Z_2|\underline{\mathcal{X}}\underline{\mathcal{Y}}} = \mathbb{W}_{Z_2|Y_2}$ , where

$$\mathbb{W}_{Z_0|X_0 Y_0}(z_0|x_0, y_0) = \begin{cases} 1 & \text{if } z_0 = x_0 = y_0 \\ 1 & \text{if } x_0 \neq y_0, z_0 = *, \\ 0 & \text{otherwise.} \end{cases}$$

The capacities of the PTP channels  $(\mathcal{X}_1, \mathcal{Z}_1, \mathbb{W}_{Z_1|X_1})$  and  $(\mathcal{Y}_2, \mathcal{Z}_2, \mathbb{W}_{Z_2|Y_2})$  is  $h_b(\frac{2}{k}) + \frac{5}{k} \log a$ .

Let  $a, k$  be chosen sufficiently/quiet large. Clearly, the shared channel  $\mathbb{W}_{Z_0|X_0 Y_0}$  (of max capacity  $\log a$ ) will be communicating most of the information, and we restrict attention to this channel through most of the following discussion. It can be verified that  $H(\underline{S})$  is close to and  $\geq \log a$  units. To communicate  $\log a$  units on  $\mathbb{W}_{Z_0|X_0 Y_0}$ ,  $X_0$  must equal  $Y_0$  whp and moreover  $X_0 = Y_0$  must be uniform.  $\underline{S}$  in Ex. 1 does not possess a GK common part, but  $S_1 = S_2$  whp. If we constrain ourselves to a S-L scheme, the distributed nature of the encoders constrains us to the S-L LMC  $X_0 - S_1 - S_2 - Y_0$ . As a consequence, our best bet at having  $X_0 = Y_0$  is to choose  $X_0 = g_1(S_1)$  and  $Y_0 = g_2(S_2)$  to be deterministic

fns of  $S_1, S_2$  respectively.<sup>3</sup> Recall that we desire  $X_0, Y_0$  to be as uniform as possible.  $\mathbb{W}_{\underline{S}}$  is so chosen such that, for large  $a, k$ , even by pooling all of the less likely symbols, the resulting pmf is quite non-uniform. A tension is apparent - randomizing  $p_{X_0|S_1}, p_{Y_0|S_2}$  can make  $X_0, Y_0$  uniform, though unequal with hp, choosing  $X_0 = g_1(S_1)$  and  $Y_0 = g_2(S_2)$ , keeps them non-uniform. Summarizing, inducing correlation onto channel inputs  $X_1, X_2$  that is constrained by the S-L LMC  $X_1 - S_1 - S_2 - X_2$  constrains our choice for  $p_{\underline{X}}$ .

We were unable to induce a desired pmf on channel inputs when constrained to SL LMC. In fact it was the  $\mathbb{W}_{\underline{S}}$  that constrained our choice of  $p_{\underline{X}}$ . If  $\underline{S}$  possessed a GK common part  $K$  with entropy  $\alpha$ , then, irrespective of  $K$ 's pmf, we could build a common (random) code with symbols iid wrt any pmf, constrained only by the entropy to  $\alpha$ . This indicates that the GK common part provides considerable flexibility when present. How do we attempt to extract such flexibility in the absence of GK common part? When  $S_1 = S_2$  whp, using inner codes of fixed block-length, we can extract a near GK common part and exploit the same. Indeed, Dueck's coding scheme accomplishes this and we build on it.

### III. COORDINATING IN THE ABSENCE OF A GK PART

In this section, we present a simplified version of the proposed coding scheme that admits a lucid description of the key elements. We derive the corresponding set of sufficient conditions and identify an example (Example 1) to prove that these conditions can be strictly less binding than the CES conditions. We begin with the main result of this section.

*Theorem 1:* A pair of sources  $(\underline{S}, \mathbb{W}_{\underline{S}})$  is transmissible over a MAC  $(\mathcal{X}, \mathcal{Y}, \mathbb{W}_{Y|\underline{X}})$  if there exists (i) a finite set  $\mathcal{K}$ , maps  $f_j : S_j \rightarrow \mathcal{K}$ , with  $K_j = f_j(S_j)$  for  $j = 1, 2$ , (ii) finite set  $\mathcal{U}$  and pmf  $p_U p_{X_1|U} p_{X_2|U}$  defined on  $\mathcal{U} \times \mathcal{X}$ , (iii)  $\delta > 0, l \in \mathbb{N}, \rho \in [0, 1]$ , (iv)  $A, B \geq 0$  such that  $\phi \in [0, 0.5]$ , where

$$\begin{aligned} \phi &:= g_{\rho, l} + \xi^{[l]} + \tau_{l, \delta}, \xi^{[l]} := 1 - (1 - \xi)^l, \\ \xi &= P(K_1 \neq K_2), \tau_{l, \delta} := 2|\mathcal{K}| \exp\{-2\delta^2 p_{K_1}^*(a^*)l\}, \\ g_{\rho, l} &= \left[ \exp\{\rho A\} \sum_{y \in \mathcal{Y}} \left\{ \sum_{u \in \mathcal{U}} p_U(u) p_{Y|U}^{\frac{1}{1+\rho}}(y|u) \right\}^{1+\rho} \right]^l, \\ A + B &\geq (1 + \delta)H(K_1), \text{ and for } j \in [2], \\ H(S_j|S_{\bar{j}}, \underline{K}) + \theta &< I(X_j; Y, X_{\bar{j}}|U) - h_b(\phi) - \phi \log |\mathcal{U}| \\ &\quad - |\mathcal{X}_{\bar{j}}||\mathcal{Y}||\mathcal{U}|(1 + |\mathcal{X}_{\bar{j}}|)\phi \log \frac{1}{\phi}, \quad (1) \\ B + H(\underline{S}|\underline{K}) + \theta &< H(X_1|U) + H(X_2|U) - H(\underline{X}|Y, U) \\ &\quad - 2h_b(\phi) - 2\phi \log |\mathcal{U}| - |\mathcal{X}||\mathcal{Y}||\mathcal{U}|\phi \log \frac{1}{\phi}, \quad (2) \end{aligned}$$

where  $\theta = h_b(\phi) + 2\phi \log |\mathcal{K}|$ .

*Remark 1:* Suppose  $\xi = 0$ , then  $K_1 = K_2$  is a GK common part. In this case, one can choose  $l \in \mathbb{N}$  arbitrarily large,  $\delta$  arbitrarily small,  $A$  smaller than, but arbitrarily close to  $I(U; Y)$ ,  $\rho$  appropriately, so that  $g_{\rho, l}, \phi, \theta$  is made as small as desired. We recover the simple set of bounds  $I(U; Y) + B >$

$H(K_1), I(X_j; Y|U, X_{\bar{j}}) > H(S_j|K_1, S_{\bar{j}}) : j = 1, 2$  and  $I(\underline{X}; Y) > B + H(\underline{S}|K_1)$ . Clearly, this is achievable. Let us briefly recollect a coding scheme for the case  $B = 0$ . Build a code  $C_U$  of rate  $I(U; Y)$  over  $\mathcal{U}$ , a 1 : 1 mapping  $e_U : T_{\delta}^l(K) \rightarrow C_U$ , and share these amongst all terminals. The encoders agree upon the chosen  $C_U$ -codeword. The rest of the information ( $S_1, S_2$  conditioned on  $K_1$ ) can be communicated assuming the knowledge of common information  $K_1, U$  at all terminals. In particular, a separation based approach involving Slepian-Wolf binning (assuming decoder side information  $K$ ) and MAC coding (conditioned on  $U$ ) does the job. Note that the block length  $l$  has to be chosen, sufficiently large, as a function of the desired probability of error.

As a consequence of the latter fact, the above scheme does not work when  $\xi$  is positive, however small. Since  $P(K_1^l \neq K_2^l) = \xi^{[l]} \rightarrow 1$  as  $l \rightarrow \infty$ , the encoders cannot agree on the chosen  $C_U$ -codeword.

*Proof:* We provide a (detailed) sketch of proof. We begin with a high level description of the coding scheme for the case  $B = 0$ . The coding scheme is similar in spirit to that described in Remark 1 except for the following key difference. The block length of  $C_U$ , and the associated mapping  $e_U$ , is fixed to  $l$ , irrespective of the desired probability of error. These  $C_U$ -codeword blocks and corresponding source blocks of length  $l$  are henceforth referred to as *sub-blocks*. Each encoder  $j$  indexes into  $T_{\delta}^l(K_1)$  and applies the same map  $e_U : T_{\delta}^l(K_1) \rightarrow C_U$ .<sup>4</sup> If  $K_j^l \notin T_{\delta}^l(K_1)$ , a codeword is chosen uniformly at random from  $C_U$ . This forms the *inner code*.

The probability that the  $C_U$ -codeword chosen by the two encoders, corresponding to a source sub-block, disagree is at most  $P(K_1^l \neq K_2^l) + P(K_1^l \notin T_{\delta}^l(K_1)) \leq \xi^{[l]} + \tau_{l, \delta}$ . Suppose the  $U - Y$  channel is memoryless<sup>5</sup> and  $g_{\rho, l}$  is an upper bound on the probability of the decoder decoding into the wrong  $C_U$ -codeword, then with probability atleast  $1 - \phi$ , the  $C_U$ -codeword chosen by the encoders agree *and* is correctly decoded at the receiver. If we consider an arbitrarily large number  $m$  of sub-blocks, then, on close to  $m(1 - \phi)$  of them, the encoders agree on the chosen  $C_U$ -codeword *and* the decoder decodes these correctly. This leads us to building an *outer code* operating over  $m$  of these sub-blocks to communicate the rest of the information ( $\underline{S}$  conditioned on the decoder's reconstruction of  $K_1, K_2$ ).<sup>6</sup>

The outer code, similar in spirit to Remark 1, is essentially based on superposition coding over  $C_U$  and Slepian-Wolf binning of the sources conditioned on the decoder's side information. However, owing to the fixed block-length of  $C_U$ , there are important challenges leading to key/technical modifications. In proceeding further, we represent the  $m$  sub-blocks of source, chosen  $C_U$ -codewords, decoded and reconstructed blocks using  $m \times l$  matrices. Sub-blocks  $1, \dots, m$  are stacked consecutively as the  $m$  rows of the matrix.

For  $j = 1, 2$ , let  $\mathbf{S}_j, \mathbf{K}_j$  denote the  $S_j$  and  $K_j$  source block matrices. For  $j = 1, 2$ , let  $\mathbf{U}_j, \mathbf{V} \in \mathcal{U}^{m \times l}$  be defined such that, for  $t \in [m]$ , (i)  $\mathbf{U}_j(t, 1 : l)$  be the encoder  $j$ 's chosen

<sup>4</sup>We emphasize that encoder 2 also indexes into  $T_{\delta}^l(K_1)$  and applies the same map  $e_U$ .

<sup>5</sup>By the interleaving technique, we ensure and argue/prove the same.

<sup>6</sup>The block length of the proposed coding scheme is therefore  $ml$ .

<sup>3</sup>Any randomization  $p_{X_0|S_1}$  and  $p_{Y_0|S_2}$  is equivalently  $X_0 = \hat{g}_1(S_1, V_1)$  and  $Y_0 = \hat{g}_2(S_2, V_2)$  for some appropriate functions  $\hat{g}_1, \hat{g}_2$  and independent RVs  $V_1, V_2$ , thereby reducing the odds of  $X_0 = Y_0$ .

$C_U$ -codeword, (ii)  $\mathbf{V}(t, 1 : l)$  be the  $C_U$ -codeword decoded into by the decoder, in sub-block  $t \in [m]$ . In every sub-block  $t$ , the decoder identifies  $\mathbf{G}(t, 1 : l) \in \mathcal{K}^l$  which will be referred to as its estimate of  $\mathbf{K}_j(t, 1 : l) : j = 1, 2$  and let  $\mathbf{G} \in \mathcal{K}^{m \times l}$  denote the corresponding matrix.

Suppose

$$\underline{S}^{lm} \underline{K}^{lm} \underline{G}^{lm} \underline{U}^{lm} \underline{V}^{lm} \sim \prod \mathbb{W}_{\underline{S}} p_{\underline{K}|\underline{G}|\underline{S}} p_{\underline{U}|\underline{V}} \quad (3)$$

has a S-L pmf, then a standard info-theoretic coding scheme implies the conditions

$$I(X_j; Y, V | X_j) - I(X_j; U_j) > H(S_j | S_j, G) : j = 1, 2 \quad (4)$$

$$I(\underline{X}; Y, V) + I(X_1; X_2) - \sum_{j=1}^2 I(X_j; U_j) > H(\underline{S} | G) \quad (5)$$

sufficient. Indeed, a separation based approach involving a Slepian-Wolf binning (assuming decoder sideinfo.  $G$ ) and MAC coding with encoder and decoder sideinfo suffices.

The fixed block-length inner code throws up two challenges. Firstly, (3) is not guaranteed. Secondly, even if we can extract certain iid sub-vectors from the  $lm$ -length vector in (3), we do not know the corresponding S-L pmf. For example, we do not have a characterization for  $p_{G|KS}$  or  $p_{V|U}$ , and hence, the quantities in (4), (5) are unknown. We overcome the first challenge via the technique of interleaving [4], and the second by upper and lower bounding the unknown quantities.

Since the inner code operates independently and identically across sub-blocks, for every  $t \in [m]$ ,

$$\mathcal{E}(t, 1 : l) := \underline{S}(t, 1 : l) \underline{K}(t, 1 : l) \underline{U}(t, 1 : l) \sim p_{\mathcal{E}^l}$$

for some joint pmf  $p_{\mathcal{E}^l}$  (that does not necessarily factor as a product of  $l$  identical factors). Also, the  $m$  vectors  $\mathcal{E}(t, 1 : l) : t \in [m]$  are independent, i.e.,  $(\mathcal{E}(t, 1 : l) : t \in [m]) \sim \prod_{t=1}^m p_{\mathcal{E}^l}$ .<sup>7</sup> In Appendix A, we prove that if, for  $t \in [m]$ ,  $\pi_t : [l] \rightarrow [l]$  are  $m$  independently and uniformly chosen permutations of the set  $[l]$ , then the  $m$  components of  $(\mathcal{E}(t, \pi_t(i)) : t \in [m])$  are iid with pmf

$$p_{\hat{S}_1 \hat{S}_2 \hat{K}_1 \hat{K}_2 \hat{U}_1 \hat{U}_2} := \frac{1}{l} \sum_{q=1}^l p_{\mathcal{E}_q}.$$

We emphasize that the pmf of  $(\mathcal{E}(t, \pi_t(i)) : t = 1, \dots, m)$  is invariant with  $i$ .<sup>8</sup> It can also be verified that  $p_{\underline{S} \underline{K}} = \mathbb{W}_{\underline{S} \underline{K} | \underline{S}}$ . Since the symbols of  $C_U$  are chosen iid  $p_U$ , we have  $\hat{U}_j \sim p_U$ .<sup>9</sup> This suggests a natural coding technique.

Randomly and uniformly partition  $\mathcal{S}_j^m$  into  $2^{mR_j}$  bins. Build channel code  $C_{X_j}$  over  $\mathcal{X}_j$  with  $2^{mR_j}$  bins, each with  $2^{mI(X_j; U_j)}$  codewords. Identify a 1-1 correspondence between the bins partitioning  $\mathcal{S}_j^m$  and the channel code bins.<sup>10</sup> Encoder  $j$  begins by identifying the inner code sub-blocks

<sup>7</sup> $p_{\mathcal{E}^l}$  is a pmf on  $(\underline{S} \times \underline{K} \times \underline{U})^l$

<sup>8</sup>This enables us build a single coding technique for each of the  $l$  vectors, as discussed in the sequel.

<sup>9</sup>We can only make statements on these marginals, but we do not have information on the joint pmf. We do know, however,  $P(\hat{U}_1 \neq \hat{U}_2) \leq \xi^{[l]} + \tau_{\delta, l}$

<sup>10</sup>Recall  $B = 0$ . If not, one of the encoders, say  $j$ , has to build  $C_j$  with  $2^{m(R_j+B)}$  bins. The  $B$  bits used to index the bin of  $C_j$  come from the index of the  $K_j^l$  within  $T_\delta^l(K_1)$ .

and populates the matrices  $\mathbf{S}_j, \mathbf{K}_j, \mathbf{U}_j$ . Since  $m$  is chosen sufficiently large, the  $l$  vectors  $(\mathbf{U}_j(t, \pi_t(i)) : t \in [m]) : i \in [l]$  are typical wrt  $p_{\hat{U}_j} = p_U$  whp. The index of the source bin containing  $\mathbf{S}_j(t, \pi_t(i) : t = 1 \dots [m])$  is looked up. A codeword in the corresponding channel code bin that is jointly typical with  $(\mathbf{U}_j(t, \pi_t(i)) : t \in [m])$  wrt  $p_{X_j|U}$  is chosen. For  $t \in [m]$ , the  $t^{\text{th}}$  coordinate of this codeword is placed in row  $t$ , column  $\pi_t(i)$  of a matrix  $\mathbf{X}_j$  that is populated with these chosen codewords. The above is performed for each  $i = 1, \dots, l$ . Encoder  $j$  has thereby populated the  $\mathbf{X}_j$  matrix. The symbols  $\mathbf{X}(1, 1 : l) \mathbf{X}(2, 1 : l) \dots \mathbf{X}(m, 1 : l)$  are placed on the channel.

The decoder populates the received matrix  $\mathbf{Y} \in \mathcal{Y}^{m \times l}$ . It can be verified that the each  $C_U$ -codeword passes through a memoryless channel. It employs a ML decoding rule on each of the rows of  $\mathbf{Y}$  and decodes into the  $C_U$ -codebook. The decoded codewords are used to populate  $\mathbf{V} \in \mathcal{U}^{m \times l}$ . Since we have assumed  $B = 0$ , the decoder can apply  $e_{C_U}^{-1}$  on each row of  $\mathbf{V}$  and populate  $\mathbf{G} \in \mathcal{K}^{m \times l}$  which is the decoders' estimate of  $\mathbf{K}_1, \mathbf{K}_2$ . Define for  $t \in [m]$

$$\mathcal{F}(t, 1 : l) := \mathcal{E}(t, 1 : l) \mathbf{X}(t, 1 : m) \mathbf{Y}(t, 1 : m) \mathbf{V}(t, 1 : m) \mathbf{G}(t, 1 : m) \sim p_{\mathcal{F}^l} \quad (6)$$

Similarly, as before, interleaving ensures (Appendix A) that the  $m$  components of  $\mathcal{F}(t : \pi_t(i)) : t \in [m]$  are iid with pmf  $p_{\hat{S} \hat{K} \hat{U} \hat{X} \hat{Y} \hat{V} \hat{G}}$  for each  $i \in [m]$ .<sup>11</sup>

The decoder now decodes into  $C_1 \times C_2$ . For each  $i$ , the decoder looks for a pair of codewords in the pair of codes  $C_1 \times C_2$  that is jointly typical wrt  $p_{\hat{X} \hat{Y} \hat{V}}$  with the vector  $(\mathbf{V}(t, \pi_t(i)) \mathbf{Y}(t, \pi_t(i)) : t \in [m])$ . If exactly one such pair of codewords is found, the corresponding bin indices is used to index the bins partitioning  $\mathcal{S}_j^m$ . In these pair of indexed bins, the decoder looks for a pair of source sequences typical with  $(\mathbf{G}(t : \pi_t(i)) : t \in [m])$  wrt  $p_{\hat{S} | \hat{G}}$ .

We only need to argue the bounds in the theorem statement. For this, we will need to compute the parameters of the outer code, particularly, the lower bounds on the size of the source partition bins and the upper bound on the size of  $C_j : j \in [2]$ . The info-theoretic quantities involved are wrt pmf  $p_{\hat{S} \hat{K} \hat{U} \hat{X} \hat{Y} \hat{V} \hat{G}}$ . In fact, the bounds are (4), (5) with  $S_j, K_j, U_j, X_j, Y, G, V$  replaced by  $\hat{S}_j, \hat{K}_j, \hat{U}_j, \hat{X}_j, \hat{Y}, \hat{G}, \hat{V}$  with pmf stated above.

It remains to prove following source and channel coding bounds. (i)  $H(\hat{S}_j | \hat{S}_j, \hat{G}) \leq H(S_j | S_j, \underline{K}) + \theta$  for  $j \in [2]$ , and (ii)  $H(\hat{S} | \hat{G}) \leq H(\underline{S} | \underline{K}) + \theta$ . We note that  $p_{\hat{S} \hat{K} \hat{G}} = p_{\underline{S} \underline{K} \underline{G}} p_{\hat{G} | \underline{K}}$  and therefore suffices to work with the latter pmf. Defining  $E := \{K_1 = K_2 = \hat{G}\}$  and noting that  $P(E) \geq 1 - \phi \geq 0.5$ , we have  $H(\underline{K} | \hat{G}) \leq H(\underline{K}, \mathbf{1}_E | \hat{G})$ , and

$$\begin{aligned} H(\underline{K}, \mathbf{1}_E | \hat{G}) &\leq H(\mathbf{1}_E) + P(E^c) H(\underline{K} | \hat{G}, \mathbf{1}_E = 0) \\ &\leq h_b(\phi) + 2\phi \log |\mathcal{K}| = \theta. \end{aligned}$$

Since  $H(S_j | S_j, \hat{G}) \leq H(S_j, \underline{K} | S_j, \hat{G}) \leq H(\underline{K} | \hat{G}) + H(S_j | S_j, \underline{K})$ , we have  $H(S_j | S_j, \hat{K}) \leq H(S_j | S_j, \underline{K}) + \theta$ , and therefore (i). Proceeding similarly, note that  $H(\underline{S} | \hat{K}) \leq$

<sup>11</sup>As before, we only know certain marginals of this pmf. For ex.,  $p_{\hat{Y} | \hat{X}} = \mathbb{W}_{Y | X} = p_{Y | X}$ .

$H(\underline{S}, \underline{K}|\hat{K}) \leq H(\underline{K}|\hat{K}) + H(\underline{S}|\underline{K}) \leq H(\underline{S}|\underline{K}) + \theta$ .<sup>12</sup> We only need to prove the following channel coding bounds:

$$\begin{aligned} I(\hat{X}_j; \hat{Y}\hat{V}\hat{X}_j) - I(\hat{X}_j; \hat{U}_j) &> I(X_j; Y, X_j|U) - h_b(\phi) \\ &\quad - \phi \log |\mathcal{U}| - |\mathcal{U}||\mathcal{X}_j||\mathcal{Y}|(1 + |\mathcal{X}_j|)\phi \log \frac{1}{\phi} \\ I(\hat{X}_j; \hat{Y}\hat{V}) + I(\hat{X}_1; \hat{X}_2) - \sum_{j=1}^2 I(\hat{X}_j; \hat{U}_j) \\ &> I(X_1; X_2|U) + I(\underline{X}; Y|U) - 2h_b(\phi) - 2\phi \log |\mathcal{U}| \\ &\quad - (1 + |\mathcal{X}|)|\mathcal{Y}||\mathcal{U}|\phi \log \frac{1}{\phi} \end{aligned} \quad (8)$$

Please refer to Appendix B for a proof.

$g_{\rho, l}$  is an upper bound on the prob. of error derived by Gallager under the assumption that the codewords are chosen uniformly. In our case this is not true. A simple technique to overcome this is to randomly and uniformly permute the  $e_U$  mapping across the  $m$  sub-blocks. This permutation is chosen upfront and agreed upon. This will uniformize the choice of codewords.<sup>13</sup> A complete proof will be provided in a subsequent version of this manuscript. ■

*Remark 2:* (1) It is not difficult to generalize the above coding technique to subsume CES scheme. For a choice of PMF  $\prod_{j=1}^2 p_{X_j|US_j}$ , the outer code has to be binned at rate  $I(X_j; S_j, U_j)$  and the chosen codeword must be jointly typical with  $U(t : \pi_t(i)), S_j(t : \pi_t(i))$ . The CES decoding scheme must be employed while decoding from the outer code. This will be provided in a subsequent version of this manuscript.

(2) A natural question is whether the proposed coding scheme proposed subsumes Dueck's scheme? The answer is NO.<sup>14</sup> Dueck coding scheme goes an additional step and 'splits the source' finer. The source in 1 has a very high variability. In other words the variance of  $-\log W_{\underline{S}}(\underline{S})$  is very large. As a consequence, if one plots  $|T_{\delta}^l(S_1)|$  and  $P(S_1^n \notin T_{\delta}^l(S_1))$  as a function of  $\delta$ , for  $l = O(k2^k)$ , it can be observed that there is not a choice for  $\delta$  for which the former and latter is small. In other words, if we choose  $\delta$  such that the former shrinks, then the latter blows up, and vice versa. If one takes a look at the region, this implies that if  $\delta$  is chosen to shrink  $\phi$ , then  $A + B$  is forced to be high, and if  $\delta$  is chosen to keep  $(1 + \delta)H(K_1)$  small, then  $\tau_{l\delta}$  blows up, causing  $\phi$  to blow up. Dueck overcomes this by spilling the source. Dueck partitions  $\mathcal{S}_j$  as  $\mathcal{A} = \{0^k\}$  and  $\mathcal{B} = \mathcal{S}_j \setminus \{0^k\}$ . Each encoder signals to the decoder whether its source is in  $\mathcal{A}$  or  $\mathcal{B}$ . This needs only  $h_b(\frac{1}{k})$  bits. Conditioned on this information being transmitted error free, the encoders now only need to consider the  $S_j \in \mathcal{B}$ . This has very low variability and can be handled via the scheme presented above. Such a splitting of the source is permitted as Dueck's coding scheme is indeed a multi-letter

scheme. We too can incorporate this in our coding scheme and will be done in a subsequent version of this manuscript.

A simple modification of Dueck's example [2] proves that conditions stated in Theorem 1 can be strictly less binding than those of CES. In describing the following example, we employ Dueck's notation. Moreover, throughout example 1, Lemma 1 and its proof, we let  $\underline{X} = X_0X_1, \underline{Y} = Y_0Y_2, \underline{Z} = Z_0Z_1Z_2$ .

*Lemma 1:* Consider Example 1 with  $\eta \geq 6$ . (i) The source and channel defined in Example 1 do not satisfy CES conditions [1, Thm 1.]. (ii) The source  $(\underline{S}, \mathbb{W}_{\underline{S}})$  is transmissible over the MAC  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathbb{W}_{\underline{Z}|\underline{X}\underline{Y}})$ .

*Proof:* (i) Following Dueck's argument [2, Sec. III.C] verbatim, we can show, for sufficiently large  $k$  and  $a \geq 2^8$

$$H(\underline{S}) \geq \frac{a^{\eta k} - 1}{ka^{\eta k}} \log(a^k - 1) \quad I(\underline{X}\underline{Y}; \underline{Z}|Q) < \frac{a^k - 1}{ka^k} \log(a^k - 1).$$

for any pmf valid  $\mathbb{W}_{\underline{S}PQP_{\underline{X}|\underline{S}Q}P_{\underline{Y}|\underline{S}Q}}\mathbb{W}_{\underline{Z}|\underline{X}\underline{Y}}$  permitted in [1, Thm. 1]. We conclude that for  $a, k$  sufficiently large, any  $\eta \geq 6$ , Example 1 do not satisfy the CES conditions.

(ii) In view of (i), it suffices to show the source is transmissible over the MAC for sufficiently large  $a, k, \eta \geq 6$ . We will come up with a choice for the parameters stated in Thm 1. For  $j \in [2]$ , set  $\mathcal{K} = \mathcal{S}_j$ ,  $f_j$  to be identity and hence  $K_j = S_j$ . Let  $\mathcal{U} = \mathcal{X}_0 = \mathcal{Y}_0$ ,  $p_{X_0X_1|U} = p_{X_0|U}p_{X_1}$ ,  $p_{Y_0Y_2|U} = p_{Y_0|U}p_{Y_2}$ ,  $U$  be uniform on  $\mathcal{U}$ ,  $X_0 = U = Y_0$ ,  $p_{X_1}$  and  $p_{Y_2}$  be capacity achieving for their respective  $\mathbb{W}_{Z_1|X_1}$  and  $\mathbb{W}_{Z_2|Y_2}$  channels. With this choice, it maybe verified that  $I(\underline{X}; \underline{Z}, \underline{Y}|U) = I(X_1; Z_1)$ ,  $I(\underline{Y}; \underline{Z}, \underline{X}|U) = I(Y_2; Z_2)$  and these quantities equal  $\mathcal{C} := h_b(\frac{2}{k}) + \frac{5}{k} \log a$  as specified. Similarly  $I(\underline{X}\underline{Y}; \underline{Z}|U) + I(\underline{X}; \underline{Y}|U) = 2\mathcal{C}$ . Moreover,  $H(S_j|\underline{K}, \mathcal{S}_j) = 0$  for  $j = 1, 2$ . It suffices to show

$$\theta + h_b(\phi) + \phi \log |\mathcal{U}| + 2|\mathcal{U}||\mathcal{Y}||\mathcal{Z}||\mathcal{X}|\phi \log \frac{1}{\phi} \leq \mathcal{C} \quad (9)$$

$$B + \theta + 2[h_b(\phi) + \phi \log |\mathcal{U}| + |\mathcal{U}||\mathcal{Y}||\mathcal{Z}||\mathcal{X}|\phi \log \frac{1}{\phi}] \leq 2\mathcal{C} \quad (10)$$

where  $\theta = h_b(\phi) + 2\phi \log |\mathcal{K}| = h_b(\phi) + 2k\phi \log a$ . Set  $\delta = \frac{1}{k}$ ,  $l = k^4 a^{\lfloor \frac{\eta}{2} \rfloor k}$ ,  $\rho = 1$ ,  $A = (1 - \frac{1}{k^3}) \log a$ ,

$$B = \left[ \left(1 + \frac{1}{k}\right) h_b\left(\frac{1}{k}\right) + \frac{1}{k} \log 2 + \frac{\eta}{a^{\eta k}} \log a \right] + \left(\frac{1}{k^3} + \frac{1}{k}\right) \log a.$$

Also note that  $\xi = \frac{1}{ka^{\eta k}}$ . The following can be verified. Firstly,

$$H(K_1) \leq H(\underline{S}) \leq h_b\left(\frac{1}{k}\right) + \log(2^{\frac{1}{k}} a) + \frac{\eta}{a^{\eta k}} \log a$$

and hence  $(A + B) \geq (1 + \delta)H(K_1) = (1 + \frac{1}{k})H(K_1)$ . Secondly,

$$\begin{aligned} \xi^{[l]} &= 1 - (1 - \xi)^l = 1 - \left(1 - \frac{1}{ka^{\eta k}}\right)^l \leq 1 - \left(1 - \frac{l}{ka^{\eta k}}\right) \\ &\leq \frac{l}{ka^{\eta k}} \leq \frac{k^3}{ka^{\frac{\eta}{2}k}}, \end{aligned}$$

where we used  $(1 - x)^n \geq 1 - nx$  for  $n \in \mathbb{N}$  and  $x \in [0, 1]$ . Thirdly,

$$\tau_{l, \delta} \leq 2a^k \exp \left\{ -2\delta^2 \frac{l}{4k^2 a^{2k}} \right\} \leq \frac{2a^k}{\exp\{\frac{1}{2}a^{(\lfloor \frac{\eta}{2} \rfloor - 2)k}\}}.$$

<sup>12</sup>Since part of the bits used to index the typical set  $|T_{\delta}^l(K_1)|$  can be communicated by one of the encoders using the outer code, the term  $B$  appears in LHS of (2).

<sup>13</sup>We do not appeal to the expurgation technique as the RV  $\hat{U}_j$  resulting expurgated codebook is not guaranteed to have marginal  $p_U$ .

<sup>14</sup>As a consequence, this work must be viewed as a first step towards a generalization that will subsume CES and Dueck's coding scheme. Further steps in this generalization will be provided in a subsequent version of this manuscript.

Fourthly,

$$g_{\rho,l} = a^{-ka^{\frac{1}{2}k}} \text{ and hence } \phi = \xi^{[l]} + \tau_{l,\delta} + g_{\rho,l} \leq \frac{2k^3}{a^{\frac{1}{2}k}},$$

for sufficiently large  $a, k$ . In order to compute the terms in (9), (10), we need to know the cardinalities of  $\mathcal{X}_1, \mathcal{Y}_2, \mathcal{Z}_1, \mathcal{Z}_2$ . Since part (i) holds for any choice of these cardinalities, we can choose these sets to have cardinality  $4a^{\frac{10}{k}}$ . By computing all the terms in (9) and (10), it can be verified that, except for  $B$  and  $C$ , the rest of the terms fall exponentially with  $k$ , for sufficiently large  $a$ . (This is because  $\phi \leq \frac{2k^3}{a^{\frac{1}{2}k}}$  falls exponentially with  $k$ .) Note that this verifies (9). In order to prove (10) it suffices to restrict attention to  $B$  and  $C$ . In fact, it suffices to show that  $2C - B$  is positive and falls polynomially with  $k$  for sufficiently large  $a$ . We can then choose a sufficiently large  $k$  for which the  $2C - B$  is positive and larger than the sum of the rest of the terms in (10). A closer look at  $B$  (11) and  $2C$  indicates that we only need to prove  $2h_b(\frac{2}{k}) - (1 + \frac{1}{k})h_b(\frac{1}{k})$  falls polynomially with  $k$ . For  $\phi \in [0, 0.5]$ , we have  $\phi \log \frac{1}{\phi} \leq h_b(\phi) \leq 2\phi \log \frac{1}{\phi}$ , and hence we have

$$\begin{aligned} 2h_b\left(\frac{2}{k}\right) - \left(1 + \frac{1}{k}\right)h_b\left(\frac{1}{k}\right) &\geq \frac{4}{k} \log \frac{k}{2} - \left(1 + \frac{1}{k}\right)\frac{2}{k} \log k \\ &\geq \frac{2}{k} \left( \log k - \frac{1}{k} \log k - \log 4 \right) \\ &\geq \frac{2}{k} \left( \log k - \frac{1}{2} \log k - \frac{1}{2} \log 16 \right) \text{ for } k \geq 2 \\ &\geq \frac{1}{k} \log \frac{k}{16} \end{aligned}$$

for sufficiently large  $k$ . We are therefore done.  $\blacksquare$

#### APPENDIX A

##### INTERLEAVING RESULTS IN IID DISTRIBUTIONS

*Lemma 2:* Let  $\mathcal{A}$  be a finite set and  $p_{A^l}$  be a pmf on  $\mathcal{A}^l$ . Let  $A(1, 1 : l), A(2, 1 : l), \dots, A(m, 1 : l) \in \mathcal{A}^l$  be independent and identically distributed vectors with pmf  $p_{A^l}$ . Let  $\Theta_l$  be the set of all permutations of the set  $\{1, 2, \dots, l\}$ . Let permutations  $\pi_1, \pi_2, \dots, \pi_m$  be chosen uniformly and independently from  $\Theta_l$ . For  $i = 1, 2, \dots, l$ , let

$$B(1 : m, i) = A(t, \pi_t(i)) : t = 1, 2, \dots, m.$$

The  $l$  vectors  $B(1 : m, i) : i = 1, 2, \dots, l$  are identically distributed with pmf  $\prod_{t=1}^m \frac{1}{l} \sum_{i=1}^l p_{A_i}$ , where

$$p_{A_i}(a) = \sum_{a_1 \in \mathcal{A}} \dots \sum_{a_{i-1} \in \mathcal{A}} \sum_{a_{i+1} \in \mathcal{A}} \dots \sum_{a_l \in \mathcal{A}} p_{A^l}(a_1, \dots, a_{i-1}, a, a_{i+1}, \dots, a_l).$$

*Proof:* For any  $i \in [l]$ , note that

$$\begin{aligned} P(B(t, i) = a_t : t \in [m]) &= \sum_{j_1 \in [l]} \dots \sum_{j_m \in [l]} P(A(t, j_t) = a_t : t \in [m]) \\ &= \frac{1}{l^m} \sum_{j_1 \in [l]} \dots \sum_{j_m \in [l]} P(A(t, j_t) = a_t : t \in [m]) \quad (11) \end{aligned}$$

$$= \frac{1}{l^m} \sum_{j_1 \in [l]} \dots \sum_{j_m \in [l]} \prod_{t=1}^m P(A(t, j_t) = a_t) \quad (12)$$

$$\begin{aligned} &= \prod_{t=1}^m \left( \frac{1}{l} \sum_{j_t \in [l]} P(A(t, j_t) = a_t) \right) \\ &= \prod_{t=1}^m \left( \frac{1}{l} \sum_{i \in [l]} p_{A_i}(a_t) \right), \quad (13) \end{aligned}$$

where (i) (11) follows from independence of the permutations  $(\pi_1, \dots, \pi_m)$  and  $A(1 : m, 1 : l)$ , (ii) (12) follows from the independence of the vectors  $A(1, 1 : l), A(2, 1 : l), \dots, A(m, 1 : l) \in \mathcal{A}^l$ , (iii) (13) follows from  $A(1, 1 : l), A(2, 1 : l), \dots, A(m, 1 : l) \in \mathcal{A}^l$  being identically distributed, and moreover,  $p_{A_i}(a) = P(A(t, i) = a)$ .  $\blacksquare$

#### APPENDIX B

##### PROOF OF CHANNEL CODING BOUNDS (7), (8)

The coding technique ensures

$$p_{\hat{U}_j, \hat{X}_j} = p_{U, X_j}. \quad (14)$$

Indeed, the symbols of  $C_U$  are chosen iid  $p_U$ . The statement proven in Appendix A implies  $p_{\hat{U}_j} = p_{U_j}$ . Moreover, recall that, the codeword from the appropriate bin in  $C_{X_j}$ , is chosen typical wrt  $p_{X_j|U_j}$ . Simple information theoretic inequalities lead us through

$$\begin{aligned} I(\hat{X}_j; \hat{Y} \hat{V} \hat{X}_j) &= I(\hat{X}_j; \hat{Y} \hat{V} \hat{X}_j \hat{U}_j) - I(\hat{X}_j; \hat{U}_j | \hat{Y} \hat{V} \hat{X}_j) \\ &\geq I(\hat{X}_j; \hat{Y} \hat{U}_j \hat{X}_j) - H(\hat{U}_j | \hat{X}_j \hat{Y} \hat{V}) \\ &\geq I(\hat{X}_j; \hat{Y} \hat{U}_j \hat{X}_j) - H(\hat{U}_j, \mathbb{1}_{\{\hat{U}_j = \hat{V}\}} | \hat{X}_j \hat{Y} \hat{V}) \\ &\geq I(\hat{X}_j; \hat{Y} \hat{U}_j \hat{X}_j) - h_b(\phi) - H(\hat{U}_j | \hat{X}_j \hat{Y} \hat{V}, \mathbb{1}_{\{\hat{U}_j = \hat{V}\}}) \\ &\geq I(\hat{X}_j; \hat{Y} \hat{U}_j \hat{X}_j) - h_b(\phi) - \phi \log |\mathcal{U}|, \end{aligned}$$

where the last two inequalities follow from  $P(\hat{U}_j = \hat{U}_j = \hat{V}) \geq 1 - \phi \geq \frac{1}{2}$ . We therefore have

$$\begin{aligned} I(\hat{X}_j; \hat{Y} \hat{V} \hat{X}_j) - I(\hat{X}_j; \hat{U}_j) &\geq I(\hat{X}_j; \hat{Y} \hat{U}_j \hat{X}_j) - I(\hat{X}_j; \hat{U}_j) \\ &\quad - h_b(\phi) - \phi \log |\mathcal{U}| \\ &= H(X_j | U) - H(\hat{X}_j | \hat{Y} \hat{U}_j \hat{X}_j) - h_b(\phi) - \phi \log |\mathcal{U}| \\ &= H(X_j | U) - H(\hat{X}_j, \hat{Y}, \hat{U}_j) + H(\hat{X}_j, \hat{Y}, \hat{U}_j) \\ &\quad - h_b(\phi) - \phi \log |\mathcal{U}| \quad (15) \end{aligned}$$

We now claim, for every  $(u, x_j, y) \in \mathcal{U} \times \mathcal{X}_j \times \mathcal{Y}$ ,

$$|p_{UX_jY}(u, x_j, y) - p_{\hat{U}_j \hat{X}_j \hat{Y}}(u, x_j, y)| \leq \phi. \quad (16)$$

Observe that

$$\begin{aligned}
p_{\hat{U}_j \hat{X}_j \hat{Y}}(u, x_j, y) &\geq p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u, x_j, y) \\
&\geq \sum_{x_j} p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u, x_j, x_j, y) \\
&\geq p_{\hat{U}_j \hat{U}_j}(u, u) \sum_{x_j} p_{\hat{X}_j \hat{X}_j \hat{Y} | \hat{U}_j \hat{U}_j}(x_j, x_j, y | u, u) \quad (17) \\
&= p_{\hat{U}_j \hat{U}_j}(u, u) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u), \quad (18)
\end{aligned}$$

where the last equality holds from the fact that  $\hat{X}_j - \hat{U}_j - \hat{U}_j - X_j, p_{X_j, X_j | \hat{U}_j \hat{U}_j}(x_j, x_j | u, u) = p_{X_j | \hat{U}_j}(x_j | u) p_{X_j | \hat{U}_j}(x_j | u) = p_{X_j X_j | U}(x_j, x_j | u)$  and  $p_{\hat{Y} | \hat{X}_1 \hat{X}_2} = p_{Y | X_1 X_2} = \mathbb{W}_{Y | X_1 X_2}$ . Proceeding further

$$\begin{aligned}
&p_{\hat{U}_j \hat{U}_j}(u, u) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) \\
&\geq (p_{\hat{U}_j}(u) - P(U_j \neq U_j)) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) \quad (19) \\
&\geq (p_{\hat{U}_j}(u) - \phi) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) \quad (20) \\
&= (p_U(u) - \phi) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) \\
&\geq p_{U X_j Y}(u, x_j, y) - \phi.
\end{aligned}$$

For the upper bound, observe that,

$$\begin{aligned}
p_{\hat{U}_j \hat{X}_j \hat{Y}}(u, x_j, y) &\leq \sum_{x_j} p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{X}_j \hat{Y}}(u, u, x_j, x_j, y) \\
&\quad + \sum_{x_j} \sum_{u_j \neq u} p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u_j, x_j, y) \\
&\leq \sum_{x_j} p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{X}_j \hat{Y}}(u, u, x_j, x_j, y) + \phi \\
&\leq p_{\hat{U}_j \hat{U}_j}(u, u) \sum_{x_j} p_{\hat{X}_j \hat{X}_j \hat{Y} | \hat{U}_j \hat{U}_j}(x_j, x_j, y | u, u) + \phi \\
&= p_{\hat{U}_j \hat{U}_j}(u, u) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) + \phi \quad (21) \\
&\leq p_{\hat{U}_j}(u) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) + \phi \\
&= p_U(u) \sum_{x_j} p_{X_j X_j Y | U}(x_j, x_j, y | u) + \phi \\
&= p_{U X_j Y}(u, x_j, y) + \phi,
\end{aligned}$$

where (21) follows from the same argument that took us from (17) to (18). We now show that, for every  $(u, \underline{x}, y) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Y}$ ,

$$|p_{U \underline{X} Y}(u, \underline{x}, y) - p_{\hat{U}_j \hat{X}_j \hat{Y}}(u, \underline{x}, y)| \leq \phi \text{ for } j = 1, 2. \quad (22)$$

Note that

$$\begin{aligned}
p_{\hat{U}_j \hat{X}_j \hat{Y}}(u, \underline{x}, y) &\geq p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u, \underline{x}, y) \\
&= p_{\hat{U}_j \hat{U}_j}(u, u) p_{\hat{X}_j \hat{Y} | \hat{U}_j \hat{U}_j}(\underline{x}, y | u, u) \\
&= p_{\hat{U}_j \hat{U}_j}(u, u) p_{X_j X_j Y | U}(x_j, x_j, y | u) \quad (23)
\end{aligned}$$

$$\begin{aligned}
&\geq (p_{\hat{U}_j}(u) - \phi) p_{X_j X_j Y | U}(x_j, x_j, y | u) \\
&= (p_U(u) - \phi) p_{X_j X_j Y | U}(x_j, x_j, y | u) \quad (24)
\end{aligned}$$

$$\geq p_{U \underline{X} Y}(u, \underline{x}, y) - \phi, \quad (25)$$

where (24) is obtained from the steps that got us from (19) to (20), and (23) follows from the argument used in going from (17) to (18). We now derive the upper bound. Arguments identical to the one used above enables us conclude

$$\begin{aligned}
p_{\hat{U}_j \hat{X}_j \hat{Y}}(u, \underline{x}, y) &\leq p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u, \underline{x}, y) \\
&\quad + \sum_{u_j \neq u} p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u_j, \underline{x}, y) \\
&\leq p_{\hat{U}_j \hat{U}_j \hat{X}_j \hat{Y}}(u, u, \underline{x}, y) + \phi \\
&= p_{\hat{U}_j \hat{U}_j}(u, u) p_{\hat{X}_j \hat{Y} | \hat{U}_j \hat{U}_j}(\underline{x}, y | u, u) + \phi \\
&= p_{\hat{U}_j \hat{U}_j}(u, u) p_{X_j X_j Y | U}(x_j, x_j, y | u) + \phi \\
&= p_{\hat{U}_j}(u) p_{X_j X_j Y | U}(x_j, x_j, y | u) + \phi \\
&= p_U(u) p_{X_j X_j Y | U}(x_j, x_j, y | u) + \phi \\
&= p_{U \underline{X} Y}(u, \underline{x}, y) + \phi. \quad (26)
\end{aligned}$$

Having proved (16) and (22), we prove the corresponding entropies are close through the following statement. Since

$$f(t) = \begin{cases} -t \log t & \text{if } t \in (0, 1] \\ 0 & \text{if } t = 0 \end{cases}$$

is concave in  $[0, 1]$  and  $f(0) = f(1) = 0$ , we have

$$|f(a) - f(b)| \leq f(|a - b|) \quad (27)$$

whenever  $0 \leq a, b \leq 1$  and  $|a - b| \leq 0.5$  [12, Proof of Lemma 2.7]. This leads us to

$$\begin{aligned}
|H(\hat{X}_j, \hat{Y}, \hat{U}_j) - H(U, \underline{X}, Y)| &\leq |\mathcal{U}| |\mathcal{X}| |\mathcal{Y}| \phi \log \frac{1}{\phi} \\
|H(\hat{X}_j, \hat{Y}, \hat{U}_j) - H(X_j, Y, U)| &\leq |\mathcal{U}| |\mathcal{X}_j| |\mathcal{Y}| \phi \log \frac{1}{\phi}.
\end{aligned}$$

Substituting the above bounds in (15), we have

$$\begin{aligned}
I(\hat{X}_j; \hat{Y} \hat{V} \hat{X}_j) - I(\hat{X}_j; \hat{U}_j) &\geq \\
&H(X_j | U) - H(\underline{X}, Y, U) + H(X_j, Y, U) - h_b(\phi) \\
&\quad - \phi \log |\mathcal{U}| - |\mathcal{U}| |\mathcal{X}_j| |\mathcal{Y}| (1 + |\mathcal{X}_j|) \phi \log \frac{1}{\phi} \\
&= I(X_j; Y X_j | U) - |\mathcal{U}| |\mathcal{X}_j| |\mathcal{Y}| (1 + |\mathcal{X}_j|) \phi \log \frac{1}{\phi} \\
&\quad - h_b(\phi) - \phi \log |\mathcal{U}|,
\end{aligned}$$

and this proves (7). We now prove (8). Observe that

$$\begin{aligned}
I(\hat{X}; \hat{Y} \hat{V}) + I(\hat{X}_1; \hat{X}_2) - \sum_{j=1}^2 I(\hat{X}_j; \hat{U}_j) &= \\
&= H(\hat{Y}, \hat{V}) - H(\hat{X}_1, \hat{X}_2, \hat{Y}, \hat{V}) + H(X_1 | U) + H(X_2 | U) \\
&= H(\hat{Y}, \hat{U}_1, \hat{V}) - H(\hat{X} \hat{Y} \hat{V} \hat{U}_1) - H(\hat{U}_1 | \hat{Y}, \hat{V}) \\
&\quad + H(\hat{U}_1 | \hat{X} \hat{Y} \hat{V}) + H(X_1 | U) + H(X_2 | U) \\
&\geq H(\hat{Y}, \hat{U}_1) - H(\hat{X} \hat{Y} \hat{U}_1) - 2H(\hat{V} | \hat{U}_1) \\
&\quad + H(X_1 | U) + H(X_2 | U) \\
&\geq H(\hat{Y}, \hat{U}_1) - H(\hat{X} \hat{Y} \hat{U}_1) - 2h_b(\phi) - 2\phi \log |\mathcal{U}| \\
&\quad + H(X_1 | U) + H(X_2 | U) \quad (28) \\
&\geq H(X_1 | U) + H(X_2 | U) - H(\underline{X} | Y, U) - 2h_b(\phi) \\
&\quad - 2\phi \log |\mathcal{U}| - |\mathcal{Y}| |\mathcal{U}| (1 + |\mathcal{X}|) \phi \log \frac{1}{\phi} \quad (29)
\end{aligned}$$



where we used  $H(\hat{V}|\hat{U}_1) \leq h_b(\phi) + \phi \log |\mathcal{U}|$  in (28). This follows from the following.

$$\begin{aligned} H(\hat{V}|\hat{U}_1) &\leq H(\hat{V}, \mathbb{1}_{\{\hat{V} \neq \hat{U}_1\}}|\hat{U}_1) \\ &\leq h_b(\mathbb{1}_{\{\hat{V} \neq \hat{U}_1\}}) + P(\hat{V} \neq \hat{U}_1)H(\hat{V}|\hat{U}_1, \mathbb{1}_{\{\hat{V} \neq \hat{U}_1\}} = 1) \\ &\leq h_b(\phi) + \phi \log |\mathcal{U}|. \end{aligned}$$

In arriving at (29), we used

$$\begin{aligned} |H(\hat{Y}, \hat{U}_1) - H(Y, U)| &\leq |\mathcal{Y}||\mathcal{U}|\phi \log \frac{1}{\phi} \\ |H(\hat{X}, \hat{Y}, \hat{U}_1) - H(X, Y, U)| &\leq |\mathcal{Y}||\mathcal{U}||\mathcal{X}|\phi \log \frac{1}{\phi}. \end{aligned}$$

The latter inequality follows from (22) and (27). In order to prove the former, it suffices to prove

$$|p_{\hat{U}_1 \hat{Y}}(u, y) - p_{UY}(u, y)| \leq \phi$$

which can be proved by following steps similar to the one we used to prove (16).

#### ACKNOWLEDGEMENTS

The author is thankful to Sandeep Pradhan and Farhad Shirani of Univ. of Michigan for discussions and sharing their insights on their findings [4], [11]. The author is thankful to Prof. Venkat Anantharam of Univ. of California at Berkeley for useful discussions on [2]. The author is thankful to Prof. P R Kumar at Texas A & M Univ for all the support, encouragement and technical inputs. The author is also thankful to Deepanshu Vasal of Univ. of Michigan for several technical discussions.

This work was supported by the Center for Science of Information at Purdue University, and in particular NSF grants NSF CCF-0939370 and NSF CCF-1140489.

#### REFERENCES

- [1] T. Cover, A. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *Trans. on Info. Th.*, vol. 26, no. 6, pp. 648 – 657, nov 1980.
- [2] G. Dueck, "A note on the multiple access channel with correlated sources (corresp.)," *Trans. on Info. Th.*, vol. 27, no. 2, pp. 232 – 235, mar 1981.
- [3] A. Wagner, B. Kelly, and Y. Altug, "Distributed rate-distortion with common components," *IEEE Trans. Info. Th.*, vol. 57, no. 7, pp. 4035–4057.
- [4] F. Shirani and S. Pradhan, "Finite block-length gains in distributed source coding," in *Proc. IEEE 2014 ISIT*, June 2014, pp. 1702–1706.
- [5] A. Lapidoth and S. Tinguely, "Sending a bivariate gaussian over a gaussian mac," *Trans. on Info. Th.*, vol. 56, no. 6, pp. 2714–2752, 2010.
- [6] C. Tian, S. Diggavi, and S. Shamai, "The achievable distortion region of sending a bivariate gaussian source on the gaussian broadcast channel," *Trans. on Info. Th.*, vol. 57, no. 10, pp. 6419–6427, Oct 2011.
- [7] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *Trans. on Info. Th.*, vol. 61, no. 4, pp. 1509–1523, April 2015.
- [8] Y. Murin, R. Dabora, and D. Gunduz, "On joint source-channel coding for correlated sources over multiple-access relay channels," *Trans. on Info. Th.*, vol. 60, no. 10, pp. 6231–6253, Oct 2014.
- [9] W. Liu and B. Chen, "Interference channels with arbitrarily correlated sources," *Trans. on Info. Th.*, vol. 57, no. 12, pp. 8027–8037, Dec 2011.
- [10] C. Tian, J. Chen, S. Diggavi, and S. Shamai, "Matched multiuser gaussian source-channel communications via uncoded schemes," in *Proc. IEEE 2015 ISIT*, June 2015, pp. 476–480.
- [11] F. Chaharsoghi, A. Sahebi, and S. Pradhan, "Distributed source coding in absence of common components," in *Proc. IEEE 2013 ISIT*, pp. 1362–66.
- [12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Budapest: Cambridge University Press, June 2011.